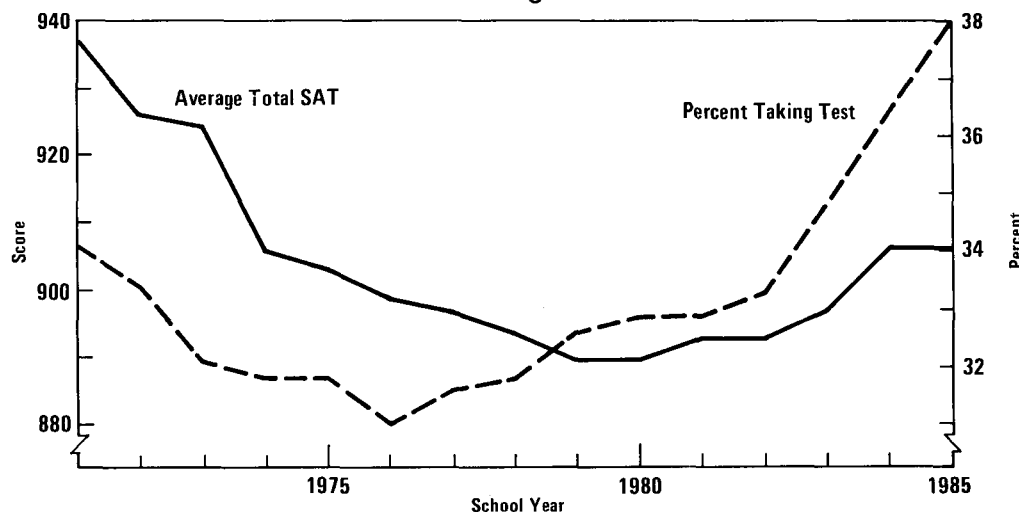


HAVE HIGH SCHOOL TEST SCORES STOPPED RISING?

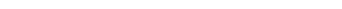
In the 1985 school year, average SAT scores remained at the level of the previous year, seemingly ending an erratic but appreciable rise that had been under way for half a decade. Some analysts quickly seized on this as evidence that the rise of achievement at the senior high school level had stagnated, even though no other major source of data suggests that scores have stopped rising in those grades.

A closer look at the SAT scores, however, shows that the current stability of average scores probably does not indicate that student performance has become stagnant. Beginning in the mid-1970s, the share of high school graduates taking the SAT grew sharply, from 31 percent in 1976 to 38 percent in 1985. Just as a similar growth in the test-taking group exacerbated the SAT decline in the 1960s, the current increase probably impeded the rise in SAT scores substantially. That is, as the pool of test-takers grows, it generally also becomes less selective, and the addition of lower-scoring students depresses average scores. If the proportion of graduates taking the test had remained constant, SAT scores would have been a better gauge of changes in student performance—but they also probably would have risen more, and 1985 scores might well have been higher than those of 1984.

The SAT: Average Scores and the Percent of Graduates Taking the Test



SOURCES: Congressional Budget Office calculations based on The College Entrance Examination Board, *National College-Bound Seniors* (New York: The College Board, various years); Office of Educational Research and Improvement, *Digest of Education Statistics, 1987* (Washington, D.C.: Department of Education, 1987); and Office of Educational Research and Improvement, unpublished data.



CHAPTER III

APPROACHES TO EXPLAINING ACHIEVEMENT TRENDS

One can easily enough devise plausible explanations of recent trends in educational achievement. The quantity and diversity of explanations that have been advanced to date give ample evidence of that. Indeed, many of the common explanations seem so persuasive that they have been subjected to relatively little scrutiny, even when they provide the rationale for formulating policy initiatives.

Yet there are many reasons to be cautious in ascribing trends to causes. Some of the common and influential explanations turn out on closer examination to be wrong; others cannot be tested with existing data. Still others appear plausible but could account for only a very small share or some particular aspect of the total movement of average test scores. Moreover, even when a factor could plausibly have contributed to the trends, the evidence typically affords much less certainty about its effects than many observers had expected.

Some attempts to explain the trends have gone amiss because they failed to distinguish between a factor's contributions to the trends and its effects on achievement more generally, and the conclusions of this analysis could likewise be misinterpreted if this distinction is not borne in mind. If this study examined only the factors' effects on achievement more generally, the methods used would be simpler, and the conclusions would in some instances be significantly different.

One approach to explaining the achievement trends is to evaluate the evidence pertaining to individual causal factors. Does the evidence suggest that changes in textbooks, for example, indeed contributed to the trends? To what aspects of the trends might they have contributed, and how big might their effects have been? By considering many diverse factors, one can gradually develop from these pieces a general view of the trends' causes. This factor-by-factor approach has characterized much of the debate to date. But many of the assessments have been incomplete, and few analysts have tried to piece the various conclusions together into a general view of the trends' possible causes.



A complementary approach, also used in this analysis, starts with the whole rather than with specific causal factors. Given the broad patterns of the achievement trends, what can one infer about likely causes? For example, one can reach different sorts of explanations on the basis of variation or lack of variation in trends among private and public schools, types of communities, age groups, students of different ability, and so on. This approach has been taken less often, perhaps because information about the broad outlines of recent trends was limited until quite recently.

EVALUATING EVIDENCE ABOUT SPECIFIC FACTORS

The first of these two methods--analyzing the evidence pertaining to one causal factor at a time--appears straightforward. In practice, however, the analyst must bear in mind a number of considerations.

Criteria for Evaluating the Effect of Specific Factors. Ideally, two criteria should be applied in evaluating the contributions of specific factors to achievement trends. The first criterion is whether a factor shows any relationship with achievement in cross-sectional studies--that is, whether an association exists between that factor and achievement levels at any given time. For example, among this year's high school seniors, do those who do more homework score better on achievement tests, all other things being equal? The second criterion--called temporal consistency here--is whether changes in an explanatory factor over time are consistent with trends in achievement. For example, have changes in the amount of homework done by typical students paralleled changes in average test scores?

Affirmative evidence about both cross-sectional relationships and temporal consistency is usually required to support a proposed explanation of the achievement trends; negative evidence about either criterion can be sufficient to refute it. Some key misconceptions about recent trends in achievement appear to have arisen because one or the other of these two criteria was paid too little heed.

No matter how strong the cross-sectional evidence pertaining to a given factor, the analyst must show temporal consistency in order to link it to specific trends in test scores. A factor that is shown by cross-sectional research to be a powerful influence on achievement in general can still be temporally inconsistent with specific trends in achievement and therefore incapable of having directly contributed to them. The importance of temporal consistency is perhaps clearest in cases where a factor of interest showed no change during the relevant period. If the amount of television

viewed, for example, did not change at all during the period of the trends in achievement being examined, one could conclude even without cross-sectional data that, whatever the effects of television viewing on achievement in general, the specific trends in test scores cannot be attributed to changes in viewing. By the same logic, finding that a certain factor could not have contributed to these trends because it was temporally inconsistent with them need not imply that it has no effects on achievement more generally or that it will not influence future trends in scores.

A lack of temporal consistency is not a problem, however, in the case of many common explanations of the achievement trends; in fact, they were first suggested precisely because they do show temporal consistency with some aspect of recent achievement trends--often congruity with trends in scores on a single test. The problem with many of these explanations is that temporal consistency alone provides little basis for concluding that a factor contributed to the trends in test scores. Innumerable factors can be found that show trends over time that are reasonably consistent with some particular aspect of trends in test scores, and yet many of these factors had no bearing on the achievement trends. To link these factors to the achievement trends, one needs some basis for judging them capable of influencing test scores. In some instances, the link may be so obvious that analysts feel no need to substantiate it. In most cases, however, cross-sectional evidence is required to establish the link.

Obstacles to Evaluating Specific Explanations of the Achievement Trends.
In their efforts to assess cross-sectional evidence and temporal consistency, researchers encounter a number of important obstacles.

In many instances, inadequacies of the existing data impede--or even preclude--an assessment of cross-sectional evidence or temporal consistency. Data about many potential causal factors are of poor quality or lacking altogether. Moreover, even when the potential explanatory factors themselves have been reasonably well measured, cross-sectional information may be so weak in other ways that only tentative conclusions--or no conclusions at all--about the factors' possible effects are warranted.

A particularly common problem in the research reviewed here is the omission or inadequate treatment of other variables--called confounded variables--that are associated with both the factors of interest and achievement and that might be responsible for the associations between them. For example, studies showing a relationship between the amount of homework and students' test scores tell little about the value of homework itself unless the studies isolate the impact of other characteristics of students who do a lot of homework and also score well on tests. Such factors might include

the students' aptitude, previous achievement, and motivation. Similarly, many studies of the relationship between class size and achievement fail to take into account decisions in some schools to assign low-achieving students to small classes in an effort to improve their performance.^{1/} Because a beneficial effect of smaller class sizes might be masked by the lower potential of students assigned to the smaller classes, such studies do not provide a good assessment of the independent effect of class size.

Even when the quality of existing data is not an obstacle, gauging temporal consistency may be complicated by the long duration of schooling. For example, in looking at a test that is administered to students after 11 years of schooling, such as the SAT, one has to decide which point during those 11 years to align with potential explanatory factors. Some analysts have searched for factors that were temporally consistent with the scores themselves, such as changes in various aspects of high school education. An alternative view is that the early years of schooling are important determinants of achievement in later grades. One analyst, for example, arguing that the early years of schooling are major determinants of reading ability, attributed trends in SAT scores to changes in the teaching of reading in primary grades a decade before each cohort took the SAT.^{2/}

Efforts to assess temporal consistency are also made more difficult by the complexity of the achievement trends themselves. When one considers a wide array of achievement tests, it becomes apparent that factors that have been singled out for attention because of their consistency with a single aspect of the achievement trends are inconsistent with other aspects. In some instances, the inconsistencies that are revealed make an explanation appear implausible altogether, but in other cases, they help clarify what the specific effects of a factor could have been. For example, some factors that have been offered as explanations are temporally consistent with test score trends in the higher grades but inconsistent with those in the elementary grades. Such factors could not have initiated trends that the relevant cohorts first exhibited when they were in the earlier grades, though they might have helped increase the severity of those trends in the higher grades.

-
1. Some of these studies also fail to consider the association between class size and the size and location of schools, which are in turn related to achievement. For a discussion of all these omissions, see Stephen N. Simpson, "Comments on 'Meta-Analysis of Research on Class Size and Achievement,'" *Educational Evaluation and Policy Analysis*, vol. 2 (May-June 1980), pp. 81-83.
 2. See, for example, Jeanne S. Chall, "Literacy: Trends and Explanations," *Educational Researcher*, vol. 12 (November 1983), pp. 3-8.

A final impediment to reaching firm conclusions is that available data generally show only that certain factors are associated with achievement and usually cannot demonstrate that those factors actually caused the trends. For example, the achievement decline among high school seniors in the 1970s was associated with a drop in the proportion of high school students enrolled in academic programs. This change might have contributed to the decline in seniors' test scores. Alternatively, both this change and the decline in test scores might have been the effects of still other factors, such as a drop in students' motivation or in their achievement at earlier grades. Indeed, both explanations could be correct. The available data are inadequate to disprove either of these competing interpretations of the observed association.

INFERRING CAUSES FROM GENERAL PATTERNS IN THE TEST SCORE DATA

The second approach to assessing the trends' causes--inferring them from the broad patterns in the achievement data--leads to very general conclusions. It might suggest, for example, that societal factors (such as demographic, cultural, and other noneducational factors) contributed to a certain aspect of the trends in achievement but give few clues about which societal factors might have been germane. The conclusions it yields are also more inferential and arguable than are those based on the assessment of individual factors. Nonetheless, this approach yields some of the most important conclusions about the trends' causes.

This alternative approach requires that one go beyond information from a single or even a few tests to discern the common threads and important divergences among various sources of test score data. One example, discussed in more detail in the following chapter, is the consistency or variation of achievement trends among diverse settings and subject areas. Despite important gaps, the available achievement data are abundant enough to make this approach possible.

To draw inferences of this sort, one needs to look not only for achievement patterns consistent with a given type of explanation, but also for patterns that are inconsistent with the alternative explanations. For example, consider the hypothesis that societal factors contributed to the trends in test scores. That hypothesis would gain support, not only from patterns in the data that could plausibly reflect societal factors, but also from patterns that are difficult to explain in terms of educational factors.



CHAPTER IV

THE POSSIBLE CAUSES OF THE ACHIEVEMENT TRENDS

Many explanations of recent trends in test scores attribute them to one or a few factors. Moreover, many analysts have focused on a single type of possible cause. One group of analysts--probably the largest and certainly the most influential--holds that the decline and subsequent upturn of test scores were largely the result of educational factors, many of which can be directly affected by explicit changes in educational policy. Another group places more of the responsibility on noneducational factors, some of which (demographic trends, for example) it sees as neutral and uncontrollable, and others of which (such as drug abuse) it regards as value-laden and alterable.

Although these views appeal to common sense and contain elements of truth, they are largely wrong. The available evidence suggests that the trends resulted from the confluence of many causes, both educational and noneducational, not from one or a few powerful factors. The individual contributions of those causes, when they can be estimated, appear to have ranged from very small to modest. In addition, many of the factors that have been cited with particular frequency turn out on closer examination to have played no role at all, and the importance of other factors cannot even be tested for want of appropriate data.

THREE GROUPS OF CONTRIBUTING FACTORS

The factors that plausibly could have contributed to the trends are extremely diverse. These factors can be organized into three broad categories:

- o Modifications of educational policy, conditions, and practice;
- o Changes in the selection of students to be tested--commonly called selection factors; and
- o Broad societal and cultural trends.

Educational factors include explicit modifications of educational policy, such as changing criteria for promoting students into subsequent grades, adopting easier textbooks, and changing the range of courses that secondary

school students are offered or required to take. The category also includes trends in educational practice that might go beyond those resulting from explicit policy changes, such as changes in the length or frequency of homework assignments, the extent of "teaching to the test," and teachers' expectations of their students. Other changes in the condition of the educational system, such as trends in the amount of experience or educational background of teachers, are also included in this category.

The term "selection factors" here refers to changes in which students from a group of potential test-takers--for example, which children of a given age--are tested.^{1/} Selection changes can stem from trends in enrollment, such as changes in the initial enrollment rates of various groups, retention rates (that is, the proportion of students from different groups remaining enrolled until a given age or grade), and the proportion of students from different groups who fall behind the typical grade level for their age.^{2/} Testing policy also can affect selection--for example, by determining which out-of-grade students, or which children with certain handicaps or with limited proficiency in English, are tested. Finally, one important aspect of selection--called self-selection--reflects students' decisions to take optional tests. Self-selection is primarily relevant to college admissions tests, such as the Scholastic Aptitude Test (SAT) and tests of the American College Testing Program (ACT).

The category of societal factors comprises all factors that are neither educational nor selection-related. It includes family composition, participation of mothers in the labor force, cultural factors such as students' attitudes toward educational success and career options, the ethnic composition of the student population, and environmental factors such as children's exposure to toxic substances.

The meaning of a change in test scores depends in part on which of these three categories is responsible for it. Test score trends attributable

-
1. The use of "selection factors" is much more specific than the more common but vaguer concept of "compositional changes." The latter concept includes all changes in the composition of the test-taking groups, regardless of whether they stem from selection or from changes in the makeup of the cohort from which the test-taking group is drawn. For example, a change in the ethnic composition of the test-taking group is a matter of selection if it stems from a change in the dropout rate among black students, but not if it reflects trends in the makeup of the school-age population as a whole. As explained below, the significance of resulting changes in test scores may hinge on this distinction.
 2. The proportion of students falling behind the typical grade affects the mix of students tested, because routine testing is commonly linked to grade levels rather than age.

to educational factors represent clear-cut changes in student performance; they reflect changes in the success with which specific skills are imparted to students, not changes in the characteristics of students entering school or selected for testing. In contrast, trends in test scores attributable to selection factors should rarely be construed as real changes in achievement or in the success of instruction. Rather, they are simply artifacts of changes in which students are chosen--or choose themselves--for testing.

To clarify this distinction, consider a high school that institutes a new program that reduces its dropout rate by half. If this change has no effect on the performance of students who would have remained in school in the absence of the new program, one would expect the school's average scores to decline because of the lower scores of students who otherwise would have dropped out. This apparent deterioration of scores, however, would indicate nothing other than the changed selection of students from the population of youth in that school's attendance area. Indeed, if the achievement of the students who were prevented from dropping out rose as a result of their remaining in school, the decline in average scores would actually be masking a real increase in the achievement level of the cohort as a whole.

Trends in test scores attributable to societal changes fall in between. Their meaning varies depending on the question at issue and the particular societal factors involved. A decline in scores attributable to a pervasive drop in students' motivation, for example, would generally be seen as a true decline in achievement. In contrast, the interpretation of a decline in scores stemming from changes in the ethnic composition of the entire school-age population is more ambiguous, assuming that scores within each ethnic group remain unchanged. If the question of interest is the achievement level of the cohort as a whole, such a decline represents a true change in performance. But if the concern is with the effectiveness of an educational system, many analysts would see such a decline as similar to a selection change, since it would not signify a deterioration of the educational performance of students from any given ethnic group.

THE EFFECTS OF EDUCATIONAL, SELECTION, AND SOCIETAL FACTORS

The evidence suggests that educational, selection, and societal factors all contributed, though in different ways, to the decline in test scores that occurred in the 1960s and 1970s.

The contributions of both educational and societal factors to the decline appear to have been considerable, and numerous factors in each group played a role. The separate effects of individual factors, however, were in most instances apparently small. The data are insufficient to indicate the relative importance of the two categories. The known contributions of selection factors to the decline, on the other hand, were limited to scores on certain optional tests taken by high school students--in particular, the SAT and ACT. Nonetheless, because the college admissions tests that were affected are among the tests that have commanded the greatest attention, selection factors have had a major effect on the public's perception of the decline in achievement.

It appears that both educational and societal factors contributed significantly to the subsequent rise in test scores as well. In contrast, insofar as they have been measured, selection changes have not contributed materially to the rise in scores and may have impeded it in some cases.

The Effects of Specific Factors

As noted in Chapter III, one approach to evaluating the origins of recent achievement trends is to examine the evidence relevant to specific factors that have been suggested as possible causes. More than two dozen such factors have been evaluated for this paper; a discussion of the evidence pertaining to each can be found in the Appendix. Many of the factors included here are frequently cited and have been particularly influential in shaping public perceptions about the causes of the achievement trends. Several factors that only rarely have been noted in this context are also included because their impact on test scores could be significant. A great many factors have been suggested as having contributed to recent achievement trends, however, and the subset discussed here is necessarily incomplete. The omission of other factors from this study does not imply that they were unimportant.

The factors considered here can be grouped into three categories: those that are plausible causes of some aspect of the trends, those that probably did not contribute appreciably, and those that cannot be assessed because there is insufficient evidence.

The factors that remain as plausible causes when systematic evidence is considered include educational, societal, and selection factors (see Table 1 and the Appendix). Although the relative importance of educational and societal causes cannot be determined, the contribution of the latter was

clearly considerable. Indeed, two factors that made particularly substantial contributions to the decline were societal: changes in the ethnic composition of the school-age population, and trends in family size (that is, the number of children per family and average birth order). 3/

A number of commonly cited educational factors could have contributed to certain aspects of the trends. A weakening of course content in the secondary grades might have contributed to the achievement decline and might help explain the greater severity of the drop among older children. Changes in the amount of homework might have contributed to both the decline and the subsequent upturn, at least among high school students. A drop in the proportion of teachers with little experience might have aided the upturn in scores, although an earlier decline in the average experience of teachers was probably unrelated to the decline in test scores.

Educational factors might also have contributed to the relative gains of minority students--that is, to the narrowing of the gap between their scores and those of nonminority students. Chapter 1 (the federally funded compensatory education program, formerly Title I) could account for some of the relative gains of both black and Hispanic students, but its contribution to this specific pattern is limited by the large proportion of nonminority students in the program and by the relatively small share of the student body that participates in the program. 4/ Desegregation also might have contributed to the relative gains of black students, but it could not have influenced the gains of Hispanics, for they did not become less segregated.

Societal factors that might have contributed to the trends are diverse and include some that have been prominent in the debate about educational achievement and others that have received little attention in this regard. Changes in the ethnic composition of the student body appear to have contributed significantly to the decline in scores but probably impeded the subsequent rise. Changes in family size brought about by the baby boom and baby bust, which have been the focus of considerable attention, probably contributed to both the decline and the upturn. Trends in the use of alcohol

-
3. Birth order refers to the sequence of births in a family; a first-born child has an order of one, a second-born, two, and so on. The average birth order of a cohort is simply the average order of all children born in that year.
 4. Chapter 1 could account for far less of the relative gains of minority students in the higher grades because of the much smaller number of older students served by the program and because the program's effects on test scores apparently largely erode after several years.

TABLE 1. THE EFFECTS OF VARIOUS EDUCATIONAL, SOCIETAL, AND SELECTION-RELATED FACTORS ON RECENT TRENDS IN TEST SCORES

Factor	Effect
Factors that Could Have Contributed	
<u>Educational</u>	
Teachers' Experience	May have contributed to upturn in scores but not to decline; effect cannot be quantified
Coursework	Change in content rather than number of courses probably contributed to the decline; cannot be quantified
Textbook Characteristics	Evidence limited to a few subjects; cannot be quantified <u>a/</u>
Homework	Possible small contribution to both decline and upturn
Title I/Chapter 1 Compensatory Education	Possible modest contribution to the relative gains of black and Hispanic students (compared with nonminority students), but only in the early grades; possible slight contribution to the relative gains of younger students
<u>Societal</u>	
Desegregation	Could account for a modest share of the relative gains of black students but not for the gains of Hispanic students
Ethnic Composition	Could have contributed one-tenth to one-fifth of the decline but impeded upturn <u>b/</u>

(Continued)

TABLE 1. (Continued)

Factor	Effect
Family Size (Number of children per family, average birth order)	Could have contributed modestly to both decline and upturn; best estimates range from 4 percent to 25 percent but are probably too high <u>b/</u>
Alcohol and Drug Use	Increase in use might have contributed to decline in higher grades; subsequent drop in use might have contributed to rise in scores
Environmental Lead	Reduced exposure to lead might have made a small contribution to the upturn
<u>Selection</u>	
Self-Selection	Contributed appreciably to decline on college admissions tests and might have impeded rise on those tests, but irrelevant to other tests

Factors that Probably Did Not Contribute Significantly

Educational

Teachers' Test Scores	Change after 1972 did not contribute to the decline; earlier data are not available
Teachers' Educational Attainment	Not temporally consistent with decline in test scores
State Graduation Requirements	No direct contribution to the decline after 1974; earlier data are not available
Minimum-Competency Testing	Did not help initiate the upturn; other effects are uncertain

(Continued)

TABLE 1. (Continued)

Factor	Effect
Head Start	No appreciable contribution to relative gains of black or Hispanic students (compared with nonminority students) after third grade; inconsequential contribution to relative gains of youngest students
<u>Societal</u>	
Single-Parent Households	Inconsequential contribution to the decline among young children; probably no appreciable effect among older children
Maternal Employment	Inconsistent data about relationship to achievement <u>c/</u>
Television Viewing	Amount of viewing did not parallel achievement trends
<u>Selection</u>	
Retention Changes	Little or no direct contribution after about 1968

**Factors About Which There is
Insufficient Evidence**

Educational

Other Characteristics
of Teachers (Such as
attitudes and morale)

Local Graduation
Requirements

(Continued)

TABLE 1. (Continued)

Factor	Effect
Grade Inflation	Inflation has been documented, but its effects have not
Demands for Writing <u>d/</u>	
<u>Societal</u>	
Students' Attitudes and Motivation	
<u>Selection</u>	
Other Selection Changes (In the testing of handicapped children, for example)	

SOURCE: Congressional Budget Office.

NOTE: For further explanation and documentation, see Appendix.

- a. Evidence in some subject areas indicates no effect.
- b. Estimates reflect only part of the relevant period.
- c. Cross-sectional evidence about the effects of maternal employment are inconsistent. Because future studies might resolve these inconsistencies, maternal employment could also be placed in the "insufficient evidence" category.
- d. Demands for writing might also be placed in the "probably did not contribute" category. Available systematic data do not indicate relevant changes in this factor but are too sparse to yield a firm conclusion.

and other drugs by high school students might have contributed to both the decline and the upturn; like changes in the content of secondary school coursework, this factor might help explain the greater severity of the decline in the higher grades. A pervasive decline in exposure to environmental lead--a factor that has received extensive attention as an influence on children's health and cognitive functioning but that has rarely been mentioned as a possible cause of the achievement trends--might also have contributed to the recent rise in test scores.

It appears that the contributions of these factors were generally less substantial than many observers have thought, ranging from very small to modest. Two factors whose effects can be estimated relatively well--changes in ethnic composition and family size--each can account for at most a fifth to a fourth of the total change in scores during portions of the decline. Although the contributions of some other factors cannot be estimated well, it appears likely that some had effects that were considerably smaller.

The factors whose hypothesized contributions to the trends are not supported by the data include some--both educational and noneducational--that have had broad acceptance and considerable influence in the public debate. For example, despite widespread concern about the effects of declining test scores of teachers, the documented decline occurred too late to have contributed to the deterioration of students' test scores. State graduation requirements have also been the focus of extensive attention but showed no appreciable change during the latter half of the decline in students' scores. (Both of these variables might have played some role, however, during the first half of the decline; the existing data do not extend back far enough to answer that question.)

Minimum-competency testing is another example; whatever its more recent effects on achievement in general--a contentious question that this analysis does not attempt to resolve--its implementation came too late to help initiate the upturn in achievement in the 1970s. A number of common societal hypotheses also fail to weather scrutiny. The rising proportion of students living in single-parent households could have contributed at most an inconsequential share of the overall decline in test scores in the early grades and probably no appreciable share of the much larger decline in the higher grades. Regardless of whether television viewing affects test scores in general, it could not have contributed significantly to the decline in test scores of the 1960s and 1970s, since the amount of viewing did not change consistently with the trends in test scores.

Finally, there is simply not enough systematic evidence to assess the effects of a number of other commonly cited factors. This gap in infor-